DOCUMENT RESUME

ED 404 360 TM 026 156

AUTHOR Devine, Marion; And Others
TITLE Higher Grade Examination Performance (1987-1994). Interchange No. 42.
INSTITUTION Scottish Council for Research in Education, Edinburgh.
SPONS AGENCY Scottish Office Education and Industry Dept., Edinburgh.
REPORT NO ISSN-0969-613X
PUB DATE 96
NOTE 12p.; For the full report, see TM 026 144.
AVAILABLE FROM RIU Dissemination Office, Scottish Council for Research in Education, 15 St. John Street, Edinburgh EH8 8JR, Scotland, United Kingdom.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Academic Achievement; Achievement Gains; Achievement Tests; *Biology; Change; *English; Foreign Countries; *Geography; *Mathematics; Secondary Education; Secondary School Students; *Standards
IDENTIFIERS *Scotland; Subject Content Knowledge

ABSTRACT

Performance over time of Scottish students was examined for Higher Grade examinations in biology, English, geography, and mathematics. Statistics have shown that the number of students gaining passes on these Higher Grade examinations is increasing from year to year, and the question of whether the standards for these examinations have been eroded has been raised. The study analyzed original test scripts, beginning with 1987 as the earliest year for which scripts were available for all subjects. In general, 100 scripts were used for each subject for each year; variations that caused small reductions in the actual number of scripts are noted in the sections describing each subject. The analysis supports the conclusion that there are no grounds for believing that there has been any change in the standards of performance for any given band of award (ranging from A to no award). The detailed differences that appeared in some instances were attributable to features of the examination papers in particular years, small sample sizes in some instances, changes to syllabi, and differences in the difficulty level of individual questions in certain papers. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED 404 360

THE SCOTTISH OFFICE

Education and Industry Department

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
R. LUAKE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

INTERCHANGE

No 42

Higher Grade Examination Performance (1987-1994)

Research and Intelligence Unit

ISSN 0969-613X

TM026156

BEST COPY AVAILABLE

Why *Interchange?*

Research cannot make the decisions for policy makers and others concerned with improving the quality of education in our schools and colleges. Nor can it by itself bring about change. However, it can create a better basis for decisions, by providing information and explanation about educational practice and by clarifying and challenging ideas and assumptions.

It is important that every opportunity should be taken to communicate research findings, both inside and outside The Scottish Office Education and Industry Department (SOEID). Moreover, if research is to have the greatest possible impact on policy and practice, the findings need to be presented in an accessible, interesting and attractive form to policy makers, teachers, lecturers, parents and employers.

Interchange aims to further improve the Research and Intelligence Unit's (RIU) dissemination of the findings of research funded by SOEID. We hope you will find that ***Interchange*** is long enough to give the flavour of the complexities, subtleties and limitations of a research study but concise enough to give a good feeling for the findings and in some cases to encourage you to obtain the full report.

*The **Interchange** symbol invites you to reflect and respond to an issue or question posed by the research. You may wish to raise awareness by responding to each **Interchange** before reading the adjacent section of text. Alternatively, you may prefer to read the text first then review each **Interchange** to construct a personal summary of the issues.*

Edited and produced for The Scottish Office Education and Industry Department by the RIU Dissemination Officer at The Scottish Council for Research in Education, October 1996.

Copyright © 1996, SOEID

***Interchange* may be photocopied for use within your own institution.**

A limited number of additional copies can be obtained by writing to the RIU Dissemination Officer at: SCRE, 15 St John Street, Edinburgh EH8 8JR. File copies for electronic downloading will also be made available on the RIU World Wide Web Server, accessible through Internet and JANET (http://www.ed.ac.uk/~riu).

Higher Grade Examination Performance (1987-1994)

Marion Devine, John Hall, Jacqueline Mapp, Kerry Musselbrook

The Scottish Council for Research in Education

There is a great deal of current interest in whether the attainment of pupils has changed over time. This often manifests itself as a concern over 'standards'. Do standards vary over time? Is it possible to find some way of checking if a consistent measurement has been applied to examination results over the years? In 1994, The Scottish Office asked the Scottish Council for Research in Education to undertake a research project investigating these issues. This Interchange summarises its findings.

This project looked at the performance of Scottish Examination Board Higher Grade candidates in four subjects – English, mathematics, biology and geography – between 1987 and 1994 in order to determine if there have been any significant changes in the standard of attainment required to gain a pass in these subjects.

The main conclusion is that:

- from the evidence available to the research project there are no grounds for believing that there has been any change in the standards of performance required to obtain any given band of award in the Higher Grade examinations which formed part of the study.

Some detailed differences did appear in the data, but these were ascribable to factors such as:

- specific features of the examination papers in particular years;
- small sample sizes for particular aspects of the study;
- changes to syllabi;
- differences in difficulty level of individual questions in certain papers.

The sample

The research team were given access to the Scottish Examination Board (SEB) archive. In each year the SEB retain one hundred scripts from candidates in each subject. These scripts represent performance within the 'range marks', which correspond to levels of performance in the examination. There are fourteen 'range marks', but these are amalgamated to produce five bands of award (bands A, B, C, D and 'no award').

Revised Higher Grade examinations were being introduced in the years covered by this study. In some years the 'revised' and 'unrevised' versions of the examination ran in parallel and the SEB retained equal numbers of each, up to a total of one hundred. In most subjects the project used scripts from the years 1987, 1988, 1993 and 1994. Any variations from this are noted in the subject sections.

Setting up the project

The research drew heavily on subject expertise provided by members of HMI, SEB Examination Officers and principal examiners, who assisted the research team throughout. The task was to re-mark scripts using uniform criteria for each subject, thus ensuring comparability between years. The criteria used conformed to general principles which stated that they had to:

- be recognisable to subject experts as a working synthesis of the subject as taught;
- comprise criteria which were discrete and observable from the scripts;
- achieve a balance between the need to describe detail and the need to work with fundamental aspects of the subject;
- be sufficiently robust to accommodate changes in the subject over the years covered by the project.

The criteria used in each subject were judged to meet these principles by the subject experts helping us. However, not all aspects of all subjects could be included in the analysis. Some detailed areas of content had changed over the years included in the study, thus preventing comparisons over time.

The criteria and methods used varied between subjects. Increased weight was given to the first of the general principles above at the expense of uniformity of methods. All marking was based on some form of criterion-referenced assessment.

Experienced Higher Grade markers were used, following training sessions in the application of the criteria. Where particular doubts were expressed about the reliability of marking procedures (for example in the marking of English essays, where potentially subjective judgements had to be made), then reliability tests were conducted.

Summaries of the findings for each subject follow.

Biology

Candidates' scripts from 1987 (unrevised), 1988 (unrevised), 1993 (revised) and 1994 (revised) were used. In 1993 some candidates sat the unrevised Higher, but we used only the revised Higher scripts, which reduced the sample for that year to fifty candidates.

Paper I of Higher Grade Biology is a multiple choice paper which draws on a bank of items kept by the SEB. It was not possible to use data from this paper in our analysis because item scores for individual candidates were not available; our findings relate only to candidates' performance in Paper II.

The subject specialists broadly categorised the questions in Biology as dealing with 'knowledge and understanding', 'problem solving' (similar to 'handling information') and 'practical applications' and used these skill areas as one basis for analysis. Another basis was found in the content areas covered. Common to all years were the following content areas:

- cell biology
- genetics and evolution
- control of growth and development
- regulation in biological systems
- adaptation

Two pairs of essay questions on very similar topics were set in different years. One (on 'industrial melanism') occurred in 1987 and 1994, and another (on 'photosynthesis') in 1988 and 1994. Special marking schemes were prepared by the team of subject experts to allow comparisons between years.

The findings in biology

There were some statistically significant differences in individual skill areas between years, but no consistent pattern to indicate any overall change in standards.

Some individual differences were due to relatively poor performance in the 1993 paper. However, the sample for that year was small, and these results must be treated with extreme caution. Apart from these, 'problem solving' and 'practical applications' showed no statistically significant differences between years.

In 'knowledge and understanding', candidates who were eventually awarded band A showed higher levels of performance in 1993 and 1994 than in 1987 and 1988. Conversely candidates who gained band C and below showed poorer performance in 1993 and 1994 than in 1987 and 1988. Our suggestion is that in the revised Higher Biology (1993 and 1994) the 'knowledge and understanding' questions were showing greater discriminatory power, being more demanding of 'average' candidates, but allowing the best to demonstrate the full extent of their knowledge.

The numbers of comparable questions were small in some years. Because of this, there was only one statistically significant difference in performance in the content areas to which we could give any weight. In genetics and evolution the best candidates (those who were awarded band A) showed better levels of performance in 1993 and 1994 than in 1987 and 1988.

Conclusions: Biology

- There was no evidence that there has been any overall change in standards in the years covered by the study.
- The evidence from the essays supports the view that there has remained consistency of standards at almost all bands of award.
- There are some suggestions that the very best candidates are exhibiting higher levels of performance in the later years of the sample.

The essay questions also provided some evidence for improved performance in later years. There were no statistically significant differences at any band of final award in the candidates' performances in the 'industrial melanism' essay. The 'photosynthesis' essay also showed no statistically significant differences in performance between years by candidates receiving bands B, C, D and 'no award'. However, the band A candidates performed significantly better in 1994 compared to 1988.

English

Scripts from 1987 (unrevised), 1988 (unrevised), 1993 (revised) and 1994 (revised) were used. The core elements of Higher Grade English remained consistent over the years from which the sample was drawn. However, some specialist areas, such as Media Studies, were excluded as they had changed considerably. Practical Criticism was excluded as it was an option chosen by few candidates and did not provide an adequate sample. The key areas included in the analysis of performance were:

- ***Literature*** - evidence came from reviews of personal reading, critical essays, and some extended responses in the Close Reading paper.
- ***Writing*** - evidence came from pieces of discursive and imaginative writing, and also from reports and/or summaries.
- ***Interpretation*** - evidence came from answers to short response questions in Interpretation and Close Reading papers.

The criteria used for marking 'Reports and Summaries' for example, included:

- understanding
- selection
- order and structure
- formal tone and register
- language competence

Thirty one such criteria were identified. Each was defined at eight levels of performance. A small sub-sample of scripts was re-marked by all markers to enable us to test the reliability of marking. This gave satisfactory results.

The Interpretation papers were dealt with differently. These consisted of short answer questions addressing the candidates' understanding and appreciation of given texts, subdivided into seven criteria, for each of which we could calculate a score. This gave us a total of 38 criteria.

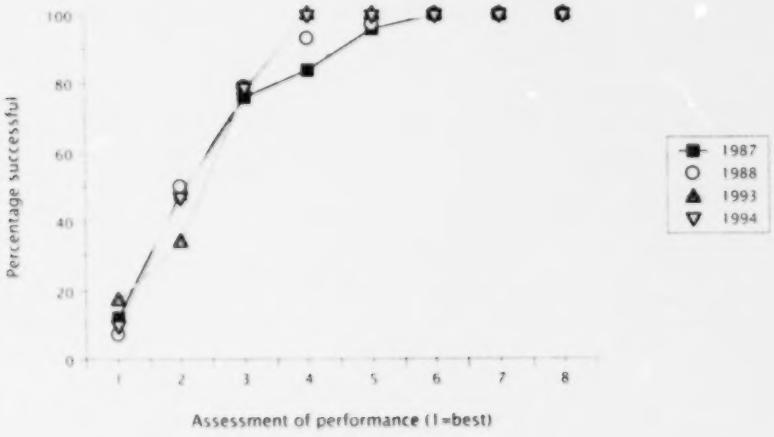
The findings in English

Overall, we found a very high degree of consistency in the performance of candidates over the years in the sample. The candidates' scripts were placed in five groups according to the final band of award they received (A to D and 'no award'). We therefore had 5 x 38 = 190 comparisons of performance in discrete

criteria. Of these, only fourteen showed any statistically significant difference in performance between years, and twelve of these differences occurred in the Interpretation paper. Typical results are illustrated graphically in Figure One which shows a cumulative frequency graph of the discursive writing 'language skills' performance of those candidates who were awarded band A. The differences between years are within the bounds of random fluctuation which could be expected with this sample.

The differences in the individual criteria in the Interpretation paper disappeared when the candidates' overall scores in Interpretation were examined. Our tentative explanation is that there are relatively small numbers of questions addressing each criterion in Interpretation each year, and that these vary in their difficulty, exaggerating apparent differences.

Figure One: Performance of band A candidates in language skills in discursive writing



Conclusions: English

- There is no evidence of any overall variation in standards between the years included in our sample.
- The few differences which existed were not indicative of any general pattern, and could be explained by features of the sample or the particular papers set in each year.

Geography

The scripts used were 1987 (unrevised), 1991 (unrevised) and 1994 (revised) examination scripts. Seven fundamental indicators of ability were selected to provide evidence for the attainment of candidates. These indicators were based on key words within examination questions which signal to the candidates what is expected of them. They cover the ability to:

- describe
- explain
- discuss
- demonstrate understanding of geographical processes
- use/demonstrate graphic skills effectively
- interpret and use OS map extracts effectively
- interpret and use effectively other resource materials provided

These indicators were accepted as fundamental for the successful study of geography.

A four point scale was chosen to assess each candidate's level of attainment on each indicator. This scale was written in fairly broad terms to indicate the qualities required to attain each level.

One area where the markers faced problems was in the assessment of the ability of candidates to use resources effectively. In some instances candidates used one fact from the diagram as a trigger and wrote extensively about that fact. In other cases, it was difficult to tell whether the candidate had made use of the resource at all.

The findings in geography

Conclusions: Geography

- Where there were significant differences in geography, these tended to be between 1987 and 1991 [both unrevised examinations] rather than between unrevised and revised examinations.
- The only significant difference in performance between unrevised and revised examinations was in the ability to use graphic skills effectively and this could be attributed to differences in the difficulty of individual questions.

Comparisons between the performance of candidates in different bands showed the same patterns in individual indicators as in the overall awards ie for each indicator, candidates with higher overall awards performed better than candidates with lower overall awards.

There were a few significant differences between the years. Most of these occurred between 1987 and 1991 (both unrevised examinations) with candidates performing better in 1987. The only significant difference between 1987 and 1994 (unrevised and revised examinations) was in the ability to use graphic skills effectively. Candidates who achieved the same band of award appeared to demonstrate higher levels of performance in 1987 than in 1994. However only a very few instances of this indicator were noted in each paper. In 1987, candidates were required to draw an annotated diagram of a standard type of settlement. In 1994, they were asked to use a diagram to help explain the formation of a river feature shown in an OS map. The question in 1987 was routine compared with that in 1994 and this was likely to be a major factor in explaining the difference.

Some concern was expressed by the markers about the lack of a sense of place or time among candidates. 'There seems to be evidence that candidates have acquired a set of broad generalisations which they apply to any place or area.' When asked to write about changes in recent years, 'answers referred back as far as the industrial revolution'. The sample of papers was small but nevertheless teachers should perhaps be aware of these concerns expressed by experienced geographers.

Mathematics

Scripts used were from 1987 (unrevised), 1988 (unrevised), 1993 (revised) and 1994 (revised). The findings from consecutive years (1987/88 and 1993/94) were combined to highlight changes in performance in unrevised and revised examinations. The greatest challenge facing the subject specialists for mathematics was the extent to which the unrevised and revised syllabuses and

examinations differed. These differences arose from the content of the syllabuses and from the approach to mathematics.

The unrevised syllabus contained many examples of mathematical content which are no longer required in the revised syllabus. This significant pruning of content in the revised Higher Grade syllabus allowed scope for more in-depth problem solving, work set in unfamiliar and 'real life' contexts and for an investigative component.

In order to find a basis for comparison, all examination papers and scripts were analysed in relation to a number of specific criteria which were largely common to both examinations. In all 180 criteria were identified. Several analyses were carried out using the criteria separately and in different combinations.

One analysis involved grouping all instances of criteria within the topics of algebra, geometry, trigonometry, differential calculus and integral calculus. As a first step in this analysis, calculations were made on the attempt rate for each group of criteria and clear differences were noted between the two sets of examinations.

The findings in mathematics

Mathematics was the subject where it was most difficult to arrive at an overall judgement about the maintenance of standards. This was partly because of changes to the syllabus, but also because of differences in the numbers of candidates not attempting some questions, or parts of questions.

For the unrevised examination, attempt rates tended to be relatively high across all candidates in all topics. In the revised examination the attempt rates for band A candidates were higher than for band D candidates. The differences ranged from 47% in algebra to 20% in geometry. Such differences might be expected in view of the structure of the revised examination. When the question is set as a problem to be solved with steps which have to be determined by the candidates, lower attaining candidates will be unable to complete full questions and hence will be unable to demonstrate attainment of those criteria contained in the latter parts. Successful candidates at Higher Grade mathematics are expected to solve problems rather than merely carry out routine procedures.

The second step of the analysis involved calculating attainment on the basis of opportunities attempted ie the results from candidates who did not attempt certain criteria were omitted from the calculation. Most of the significant differences occurred in integral calculus. These differences showed higher than statistically expected numbers of correct responses in the unrevised examinations for candidates at bands A, B and C. A number of reasons can be suggested for these differences. The criteria grouped into topic areas included criteria which were judged to be straightforward and those which were more difficult. It could be that the criteria which were judged to be more difficult were mainly from the revised examinations. A separate analysis based only on the

Conclusions: Mathematics

- When comparisons were made in relation to instances of straightforward criteria grouped under topic headings, there were no significant differences between performance in revised and unrevised examinations.
- When all instances of criteria, both straightforward and more difficult, were included there were significant differences at bands A, B and C in integral calculus and in each case there were more correct responses in the unrevised examinations. One probable explanation was that most of the difficult instances of criteria were in the revised examinations.
- In geometry there were significant differences in performance at all bands between the years and in each case there were more correct responses in the revised examinations.

straightforward criteria showed no significant differences and seemed to confirm this as a possible explanation. The completely different styles of the examination questions could also be an explanatory factor. Even where candidates can respond correctly in a series of familiar routines, they may be unable to do so in more extended problems in unfamiliar contexts.

Significant differences were also noted in geometry. The results for all bands showed higher than expected numbers of correct responses in the revised examinations.

Finally, markers also graded each candidate for his/her ability to present work clearly and in a logical manner. No significant differences were found in the performance between candidates in the different years.

Final observations

We began by noting the concern with 'standards' in education common today. We can conclude from this study that the 'standards' applied in Higher Grade examinations are being maintained, insofar as we are able to judge from the evidence available to us.

We end with a few observations for any similar exercise in the future.

This research was unable to take account of internally assessed elements of Higher Grade examinations, and of those papers which consisted of confidential item-banked questions, and some thought could be given to this in future exercises.

In this first attempt to look back at performance, we gave great emphasis to the appropriateness of the criteria used, and this led to different approaches for each subject. It may be possible in future, by giving more consideration to the type and scale of the criteria, to introduce greater commonality of approach. We would caution, however, that this would make great demands on future researchers and would require a large amount of preparatory time.

Finally we note that with the changes now being made in relation to *Higher Still* careful thought should be given to the possibilities of repeating the exercise. We now have a base-line of information on the revised examinations on which we can build.

Final report

Further details of the study are in the final report - ***Maintaining Standards: Performance at Higher Grade in Biology, English, Geography & Mathematics*** - which can be purchased from the Scottish Council for Research in Education, 15 St John Street, Edinburgh, EH8 8JR. Tel: 0131 557 2944.

The views expressed in this paper are those of the authors and do not necessarily reflect those of The Scottish Office Education and Industry Department who funded the study.

The Interchange series

1. Homework Policy and Practice
2. School to Higher Education: Bridging the Gap
3. Teaching, Learning and Assessment in the National Certificate
4. Developing School Managers
5. Transition from School to Adulthood of Young People with Recorded Special Educational Needs
6. Discipline in Scottish Schools
7. Training the Trainers' Programmes: Effective Management and Monitoring
8. Introduction of the New Further Education College Council System
9. Young People's Experience of National Certificate Modules
10. Costs and Benefits of Adult Basic Education
11. Performance Indicators and Examination Results
12. An Evaluation of the Advanced Courses Development Programme
13. Staying the Course
14. A Study of Probationer Teachers
15. Making School Boards Work
16. Professional Development Through Research
17. Students' Views on SWAP
18. Specific Learning Difficulties: Policy, Practice and Provision
19. Foreign Languages in Primary Schools: the National Pilot Projects in Scotland
20. Towards More School Based Training?
21. Patterns of Attainment in Standard Grade Mathematics 3–6
22. Patterns of Attainment in Standard Grade English 3–6
23. Implementing 5–14: a Progress Report
24. Education–Business Links: Patterns of Partnership
25. Foreign Languages for Vocational Purposes in Further and Higher Education
26. School for Skills
27. Effective Support for Learning: Themes from the RAISE Project
28. Marketing Means Business
29. Adult Education: Participation, Guidance and Progression
30. Studies of Differentiation Practices in Primary and Secondary Schools
31. Health Education: What Do Young People Want To Know?
32. Social and Educational Services for Children Under Five
33. Issues in Teachers' Continuing Professional Development
34. Primary Teachers' Understanding of Concepts in Science and Technology
35. Putting 5-14 in Place: An Overview of the Methods and Findings of the Evaluation 1991-95
36. Implementing 5-14 in Primary Schools
37. Implementing 5-14 in Secondary Schools
38. Provision for Special Educational Needs
39. Methods of Teaching Reading
40. Criteria for Opening Records of Needs
41. Guidance in Secondary Schools

Further information

If you have views on ***Interchange*** and/or wish to find out more about RIU's research programme, contact the Research and Intelligence Unit (RIU), The Scottish Office Education and Industry Department, Room 1-B, Victoria Quay, Edinburgh EH6 6QQ

Printed by GNP Booth Ltd, Glasgow